

Example 宝可梦の战斗力回归

$$y = wx_{cp} + b$$

Loss function :

描述模型表现好不好的函数。——也即衡量参数b和w的好坏。

挑选表现最好的function。

优化方法: gradient descent 只要函数可微分, 就可以应用这一方法。

其具体流程:

- 1.从任意初始参数出发
- 2.损失函数对w求偏导
- 3.w向负梯度方向移动, 乘上学习率 η

"gradient"其实就是所有参数对loss求导的结果。

模型复杂程度的提升反而可能带来过拟合的问题——泛化误差提升

理论上高阶模型能够拟合低阶模型, 但高阶模型会overfitting(过拟合), 从而导致模型性能下降。

然而, 收集60只宝可梦之后发现, 实际上模型应当将物种纳入考虑, 不同的物种有不同的模型参数。最终模型如下所示。

$$\begin{aligned}
y = & b_1 \cdot \delta(x_s = \text{Pidgey}) \\
& + w_1 \cdot \delta(x_s = \text{Pidgey})x_{cp} \\
& + b_2 \cdot \delta(x_s = \text{Weedle}) \\
& + w_2 \cdot \delta(x_s = \text{Weedle})x_{cp} \\
& + b_3 \cdot \delta(x_s = \text{Caterpie}) \\
& + w_3 \cdot \delta(x_s = \text{Caterpie})x_{cp} \\
& + b_4 \cdot \delta(x_s = \text{Eevee})
\end{aligned}$$

综合考虑：

把所有可能属性全塞进网络。

但是这样训练的模型出现了非常严重的overfitting

咋办? Regularization!

Redesign function

$$L = \sum_n \left(\hat{y}^n - \left(b + \sum w_i x_i \right) \right)^2 + \lambda \sum (w_i)^2$$

加入正则项将损失函数正则化。 λ 的加入约束函数的参数越小越好，这样确保了函数对输入的变化较为平滑(对输入的变化不那么敏感)。当 λ 越大时，找到的函数越平滑，但可能此时error会增大，二者需要得到一个平衡。

λ	Training	Testing
0	1.9	102.3
1	2.3	68.7
10	3.5	25.7
100	4.1	11.1
1000	5.6	12.8
10000	6.3	18.7
100000	8.5	26.8

平滑的函数很好，但不能太过于平滑。

宝可梦の分类问题

input = 杰尼龟, output = 水

建立一个分类模型：其实就是一个SVM支持向量机即可完成分类的任务。不能使用Regression的方法来处理！

高斯似然分布 —— 最大似然 找到似然最大的 μ 和 Σ ，其实就是所有点的几率。

然而高斯分布的区分效果也没那么好。

在共用参数 Σ 的情况下，七维空间下模型具有较好的分类表现。

算法步骤：

1. 给出一个 X ，求出对应的似然度 $P(C1 | X)$ 如大于阈值则分类为第一种

2.找到分类最精确的参数(最大化)

Q:为什么用Gaussian?

A:自己选的。

可以按照实际的情况选朴素贝叶斯，高斯，伯努利等等。

逻辑回归 多类问题的分类操作

每一类给出一组参数，丢进SoftMax function，取得到结果的最大值作为分类结论。

通过恰当的特征变换可能得到表现更好的模型。

怎样找到好的变换？

将输入经过一层layer变换转换成新的特征再分类，得到的新模型就可以拟合特征变换的步骤。

DL 的步骤，机器学习优化攻略

function, Loss, optimize

优化思路：

1.检查在training data上的loss,分析原因：

model bias : 模型过简单，重新设计，为模型赋予更大的弹性(更多的feature，更大的model)

optimize做的不好：gradient descent存在问题，可能停留在局部最优

究竟是哪一个？

判断方法：比较不同的模型(不同层数)未必是overfitting! 在训练资料上56层还是比24层弱，这说明56层模型的 optimization方法不好。因为56层一定比24层弹性强，所以一定能拟合出至少跟24层一样的模型。

2.在test上检查：可能出现overfitting 的现象：模型对于training分布的拟合过于精确，泛化能力下降。

解决办法：

1.增加训练资料

2.data augmentation :自助法

如：照片反转，信息截取(通过合理的方式进行数据处理，获得更多的数据)

3. 给模型一定的限制，如：较少的参数，共用参数(让一些参数有共同数值,CNN中会提及)，从Fully-connected到CNN, 这与模型的bias之间是矛盾的，需要权衡。

4. Regularization , Dropout

也可能出现mismatch的情形：出现反常数据，训练资料和测试资料的分布完全不同。

